

Matteo Riondato

CONTACTS	Stanford University 353 Serra Mall Stanford, CA 94306, USA	Voice: +1-401-654-3216 E-mail: rionda@stanford.edu WWW: http://cs.brown.edu/~matteo
RESEARCH INTERESTS	Data analytics, pattern mining, graph mining, randomized algorithms, social network analysis, algorithms for distributed/parallel architectures for big data, privacy issues in data analysis, statistical learning theory.	
CURRENT POSITION	Stanford University , Stanford, California, USA Postdoctoral Researcher, since August 2014 <ul style="list-style-type: none">Supervisor: Prof. Chris Ré	
EDUCATION	Brown University , Providence, Rhode Island, USA Ph.D., Computer Science, May 2014 <ul style="list-style-type: none">Dissertation: <i>Sampling-based randomized algorithms for Big Data analytics</i>Advisor: Prof. Eli UpfalBest student poster award at SIAM Int. Conf. on Data Mining (SDM'14), April 2014. M.Sc., Computer Science, May 2010 Università di Padova , Padua, Italy Laurea Magistrale (M.Sc.) 110/110 <i>cum laude</i> , Computer Engineering, July 2009 <ul style="list-style-type: none">Master Thesis: <i>Top-k frequent itemsets mining through sampling</i>Advisors: Prof. Andrea Pietracaprina, Prof. Eli Upfal, Fabio Vandin Laurea (B.Sc.), Information Engineering, July 2007 <ul style="list-style-type: none">Final Project: <i>Algorithmical foundations of cryptography</i>Advisor: Prof. Andrea Pietracaprina	
PUBLICATIONS	<p>[12] M. Riondato, D. García-Soriano, and F. Bonchi. Graph summarization with quality guarantees. <i>14th IEEE Int. Conf. on Data Mining, ICDM'14, Shenzhen, China, December 14–17, 2014</i></p> <p>[11] M. Riondato. Sampling-based data mining algorithms: Modern techniques and case studies. <i>European Conf. on Mach. Learn. and Knowl. Disc. in Databases, ECML PKDD'14, Nancy, France, September 15–19, 2014</i>.</p> <p>[10] M. Riondato and F. Vandin. Finding the true frequent itemsets. <i>14th SIAM Int. Conf. on Data Mining, SDM'14, Philadelphia, PA, USA, April 24–26, 2014</i>.</p> <p>[9] M. Riondato and E. Upfal. Efficient discovery of association rules and frequent itemsets through sampling with tight performance guarantees. <i>ACM Trans. on Knowl. Disc. from Data</i>, 8(4):20, 2014.</p> <p>[8] M. Riondato and E. M. Kornaropoulos. Fast approximation of betweenness centrality through sampling. <i>7th ACM Int. Conf. on Web Search and Data Mining, WSDM'14, New York, NY, USA, February 24–28, 2014</i>.</p> <p>[7] M. Riondato, J. A. DeBrabant, R. Fonseca, and E. Upfal. PARMA: A parallel randomized algorithm for association rules mining in MapReduce. <i>21st ACM Int. Conf. on Inform. and Knowl. Manag., CIKM'12, Maui, HI, USA, October 29 – November 02, 2012</i>.</p> <p>[6] M. Riondato and E. Upfal. Efficient discovery of association rules and frequent itemsets through sampling with tight performance guarantees. <i>European Conf. on Mach. Learn. and Knowl. Disc. in Databases, ECML PKDD'12, Bristol, UK, September 24–28, 2012</i>.</p> <p>[5] A. Pietracaprina, G. Pucci, M. Riondato, F. Silvestri, and E. Upfal. Space-round tradeoffs for MapReduce computations reduce computations. <i>ACM Int. Conf. on Supercomputing, ICS'12, Venice, Italy, June 25–29, 2012</i>.</p> <p>[4] M. Akdere, U. Çetintemel, M. Riondato, E. Upfal, and S. B. Zdonik. Learning-based query performance modeling and prediction. <i>28th IEEE Int. Conf. on Data Engin., ICDE'12, April 1–5, 2012, Washington, DC, USA</i>.</p> <p>[3] M. Riondato, M. Akdere, U. Çetintemel, S. B. Zdonik, and E. Upfal. The VC-dimension of SQL queries and selectivity estimation through sampling. <i>European Conf. on Mach. Learn. and Knowl. Disc. in Databases, ECML PKDD'11, Athens, Greece, September 5–9, 2011</i>.</p>	

[2] M. Akdere, U. Çetintemel, M. Riondato, E. Upfal, and S. B. Zdonik. The case for predictive database systems: Opportunities and challenges. *5th Biennial Conf. on Innovative Data Sys. Research, CIDR'11, Asilomar, CA, USA, January 9–12, 2011*.

[1] A. Pietracaprina, M. Riondato, E. Upfal, and F. Vandin. Mining top-k frequent itemsets through progressive sampling. *Data Mining and Knowl. Disc.*, 21(2):310–326 (special issue for ECML PKDD'10), 2010.

UNDER SUBMISSION

[13] A. Anagnostopoulos, L. Becchetti, A. Fazzone, I. Mele, and M. Riondato. The importance of being experts: Efficient max-finding in crowdsourcing.

[14] M. Riondato, M. Akdere, U. Çetintemel, S. B. Zdonik, and E. Upfal. The VC-dimension of SQL queries and selectivity estimation through sampling (extended version).

AWARDS	SIAM Best Student Poster Award at SIAM SDM'14	April 2014
	SIAM Student Travel Grant to SIAM SDM'14	April 2014
	Brown University Dissertation Fellowship	Fall 2013
	Yahoo! Research Summer Internship	Summer 2013
	Research Fellowship from MIUR of Italy under Project AlgoDeep prot. 2008TFBWL4	Summer 2011
	Brown University Graduate Fellowship	Academic Year 2009–10

INVITED TALKS	Efficient Frequent Itemsets mining through sampling <i>Database Research Group Seminar, University of Waterloo, 5/7/2014.</i>
	Efficient Frequent Itemsets mining through sampling <i>Database Group Seminar, MIT CSAIL, 04/10/2014.</i>
	Taming the challenges of Big Data with statistical data analytics. <i>Department of Computer Science, Boston College, 02/07/2014.</i>
	Fast betweenness estimation through sampling. <i>Data Management Group Seminar, Boston University, 10/17/2013.</i>
	Fast betweenness estimation through sampling. <i>Lab Research Seminar, Yahoo! Labs Barcelona, 6/13/2013.</i>
	Fast betweenness estimation through sampling. <i>Advanced Computing Group Talk, Department of Information Engineering, University of Padua, Italy, 5/30/2013.</i>
	Statistical learning theory meets knowledge discovery: Randomized algorithms for Big Data analytics. <i>Brown CS Industrial Partners Program Symposium, 4/25/2013.</i>
	Approximate aggregate database queries through sampling. <i>Invited Lecture for CSCI-2950-T, Brown University, 10/25/2011.</i>
	Graphs algorithms in MapReduce: Design choices and optimizations. <i>Invited Lecture for CSCI-2950-U, Brown University, 9/27/2011.</i>
	Statistical learning theory meets databases. <i>Advanced Computing Group Talk, Department of Information Engineering, University of Padua, Italy, 06/24/2011.</i>
	Top-k frequent itemsets mining through sampling. <i>Advanced Computing Group Talk, Department of Information Engineering, University of Padua, Italy, 9/16/2010.</i>

ADDITIONAL RESEARCH EXPERIENCE	Yahoo! Labs , Barcelona, Spain	
	<i>Summer Intern Research Scientist in the Web Mining Group</i>	June – August 2013
	<ul style="list-style-type: none"> Worked with Dr. Francesco Bonchi and others on algorithms for graph summarization and for frequent itemsets mining from data streams in a distributed setting. 	
	Summer School on Massive Data Mining , Copenhagen, Denmark	
	<i>Student</i>	August 8 – 10, 2012
	<ul style="list-style-type: none"> Attended the Summer School on Massive Data Mining organized by Prof. R. Pagh at IT University of Copenhagen. 	

	<p>Sapienza Università di Roma, Rome, Italy <i>Visiting Ph.D. Student</i> June – July 2012</p> <ul style="list-style-type: none"> • Worked with Prof. S. Leonardi, A. Anagnostopoulos, and L. Becchetti on algorithms for a model of crowdsourcing computation, and on algorithms for MapReduce. <p>Università di Padova, Padua, Italy <i>Research Fellow</i> May – July 2011</p> <ul style="list-style-type: none"> • Worked with Prof. A. Pietracaprina, G. Pucci, and F. Silvestri on “The MapReduce Paradigm: computational model and algorithms”. Funded by MIUR of Italy under Project AlgoDEEP prot. 2008TFBWL4. <p>Chalmers University of Technology, Gothenburg, Sweden <i>Visiting Ph.D. Student</i> August 2010</p> <ul style="list-style-type: none"> • Worked with Prof. Devdatt Dubhashi and helped organize a seminar on Probability and Computing. <p>Brown University, Providence, RI, USA <i>Visiting Grad Student</i> October 2008 – June 2009</p> <ul style="list-style-type: none"> • Worked with Prof. Eli Upfal on master thesis <i>Top-K Frequent Itemsets Mining through Sampling</i>.
TEACHING AND ADVISING TRAINING AND EXPERIENCE	<p>The Harriet W. Sheridan Center for Teaching and Learning (Brown University), Providence, RI, USA</p> <ul style="list-style-type: none"> • Teaching Certificate I: Reflective Learning, AY 2013–14 <p>Brown University, Providence, Rhode Island, USA</p> <ul style="list-style-type: none"> • Mentor, New Scientist Program Graduate-Undergraduate Mentoring Initiative, Spring 2014 • Teaching Assistant, Probability and Computing, Spring 2012, Spring 2013, Spring 2014 • Teaching Assistant, Probabilistic Methods in Computer Science, Fall 2010
SERVICE	<p>PC member for CIKM’14 (Knowledge Management track)</p> <p>Reviewer for IEEE TKDE</p> <p>Reviewer for the following conferences</p> <ul style="list-style-type: none"> • RANDOM’11, ICS’12, IPDPS’12, WSDM’13, MFCS’13, BigData’13, WSDM’14, DISC’14. <p>Brown University Graduate Student Council <i>President</i> April 2011 – December 2012</p> <ul style="list-style-type: none"> • Represented the interests of the entire graduate student community (approx. 2000 students) with the university administration and the broader community at all levels. Previously held positions include Vice-President of Administration (Jan. – April 2011) and representative for the Computer Science Department (Sept. 2009 – April 2011, Jan 2013 – <i>ongoing</i>). <p>Brown University Graduate Council <i>Student Representative</i> September 2011 – May 2013</p> <ul style="list-style-type: none"> • Represented the graduate student community in the highest body governing graduate education. <p>Brown University Strategic Planning Committee on Doctoral Education <i>Student Representative</i> September 2012 – May 2013</p> <ul style="list-style-type: none"> • Represented the graduate student community to develop the presidential strategic plan for the next decade. <p>Brown Computer Science Theory Lunch <i>Organizer</i> January – December 2010</p> <ul style="list-style-type: none"> • Responsible for organizing the weekly meeting of the Theory Group.
MEMBERSHIP	<p>Association for Computing Machinery (ACM)</p> <ul style="list-style-type: none"> • Student member and member of Special Interest Group on Knowledge Discovery from Data (SIGKDD). <p>Institute of Electrical and Electronic Engineers (IEEE)</p> <ul style="list-style-type: none"> • Student member and member of IEEE Computer Society. <p>Society for Industrial and Applied Mathematics (SIAM)</p> <ul style="list-style-type: none"> • Student member and member of SIAM Activity Group on Data Mining and Analytics (SIAG/DMA).

SOFTWARE

CentrSampl

- Algorithm to estimate node betweenness centrality in large graphs. Based on [8].
<http://cs.brown.edu/~matteo/centrsampl.tar.bz2>.

PARMA

- Frequent itemsets and association rules mining algorithm for Hadoop MapReduce. Based on [7].
<http://cs.brown.edu/~matteo/parma.tar.bz2>.

FreeSBIE

Developer, Release Engineer for the 2.x series

April 2004 – July 2009

- Developed FreeSBIE, a Live-CD distribution of FreeBSD bootable from CD-ROM. Release Engineer for FreeSBIE 2.X. series, responsible for all aspects of the release. <http://www.FreeSBIE.org>.

The FreeBSD Project

src Committer

January 2006 – June 2013

- Contributed to the development of the FreeBSD UNIX operating system. Granted write access to the main source repository. Worked on the `jail` security feature and on handling and solving bug reports of various nature. Author of the *Jail* chapter in the FreeBSD Handbook. <http://www.FreeBSD.org>.

REFERENCES

Prof. Eli Upfal	Prof. Uğur Çetintemel	Dr. Francesco Bonchi
Dept. of Computer Science	Dept. of Computer Science	Yahoo! Labs Barcelona
Brown University	Brown University	Yahoo!
eli@cs.brown.edu	ugur@cs.brown.edu	bonchi@yahoo-inc.com

UP-TO-DATE VERSION

Available from http://cs.brown.edu/~matteo/matteo_riondato_cv.pdf